Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

Uncertainty-aware traffic accident risk prediction via multi-view hypergraph contrastive learning

Yimei Zhang [a,b], Guojiang Shen [a,b], Wenyi Zhang [a,b], Kaili Ning [a,b], Renhe Jiang [c], Xiangjie Kong [a,b],*

[a] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China
[b] Zhejiang Key Laboratory of Visual Information Intelligent Processing, Hangzhou, 310023, China
[c] Center for Spatial Information Science, The University of Tokyo, Bunkyo-Ku, 1138654, Japan

ARTICLE INFO

Keywords:
Traffic accident prediction
Multi-source data fusion
Uncertainty quantification
Spatial–temporal data

ABSTRACT

Traffic accident prediction is crucial for maintaining safety in smart cities. Accurate prediction can significantly reduce casualties and economic losses, while alleviating public concerns about urban safety. However, achieving this is challenging. First, accident data exhibits twofold imbalances: (i) a class imbalance between accident occurrence and non-occurrence, and (ii) a spatial distribution imbalance among different regions. Second, sporadic traffic accidents result in sparse supervised signals, limiting the spatial–temporal representations of conventional deep models. Lastly, the Gaussian assumption underlying the previous deterministic deep learning models is unsuitable for accident risk data characterized by dispersed and many zeros. To address these challenges, we propose an **U**ncertainty-aware spatial–temporal multi-view hypergraph contrastive learning framework for **T**raffic **a**ccident **r**isk prediction (TarU). This framework not only jointly captures local geographical spatial–temporal and global semantic dependencies from different views, but also parameterizes the probabilistic distribution of accident risk to quantify uncertainty. Particularly, a hypergraph-enhanced network and an auxiliary contrastive learning architecture are designed to enhance self-discrimination among regions. Extensive experiments on two real-world datasets demonstrate the effectiveness of TarU. The proposed framework may also be a paradigm for addressing spatial–temporal data mining tasks with sparse labels.

1. Introduction

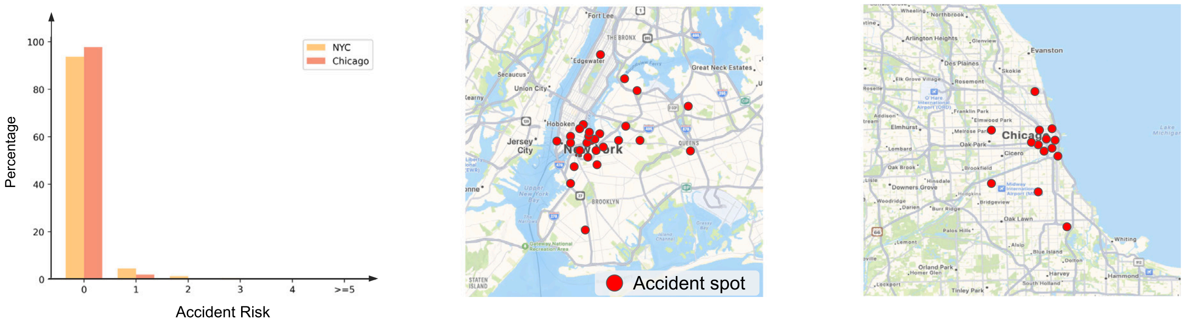
With the acceleration of urbanization, the number of vehicles on the roads has proliferated over the past few decades [1]. The significant increase in traffic accidents has emerged as a critical socio-economic challenge for humanity. According to the Global Status Report on Road Safety (GSRRS) 2023 published by WHO,[1] the number of road traffic deaths reaches 1.19 million per year. Fortunately, the report also indicates that effective measures can significantly reduce fatalities. A real-world example is that the fatality rate in Tennessee has been reduced by 8.16% in 2016 after deploying an accident prediction model [2]. Therefore, accurate traffic accident prediction is crucial for helping governments implement certain traffic controls and strategies to mitigate the harm caused by crashes.

Recent studies [3–5] attempt to obtain more accurate predictions by designing sophisticated deep learning models to analyze the factors that affect the occurrence of accidents. However, traffic accident risk prediction problem presents unique challenges. First, accident data exhibits significant imbalance, primarily in two dimensions: (1) Due to the sporadic nature of accidents, most data points are zero, with only a few showing positive values, reflecting abnormal traffic accidents, which results in the first imbalance: class imbalance between accident and non-accident cases, as seen in Fig. 1(a). Such class imbalance will induce a bias towards dominant classes, leading to zero-inflated problems where the models tend to predict all outcomes as zero. Previous work [6] sidesteps this challenge by using a lower resolution, such as a day. (2) Neural network-based models tend to be influenced by regions with high accident frequencies when learning the spatial dependencies between geographical locations. Consequently, these models tend to assign higher risk values to urban areas while overlooking relatively high-risk regions in rural areas. This causes the second imbalance: spatial distribution imbalance among regions, as shown in Fig. 1(b). However, neighborhood information aggregation mechanism in current Graph Neural Network (GNN) methods [7–11] will exacerbate this phenomenon [12,13].

* Corresponding author at: College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, 310023, China.
E-mail address: xjkong@ieee.org (X. Kong).

[1] https://www.who.int/

https://doi.org/10.1016/j.inffus.2025.103331
Received 23 December 2024; Received in revised form 14 March 2025; Accepted 13 May 2025
Available online 29 May 2025



(a) Class imbalance

(b) Spatial distribution imbalance

Fig. 1. Twofold imbalances in traffic accident risk data.

Second, the large number of zeros and asymmetric data distributions deviate from the Gaussian assumptions of the previous deep models. Distributions such as the Zero-Inflated Negative Binomial (ZINB) may be more appropriate to model highly-sparse and over-dispersed accident risk data. Moreover, an important issue in understanding transportation risk prediction is quantifying uncertainty by considering distributions rather than averages. Many processes have inherently stochastic properties, so outputting a probability distribution is closer to the essence, and the accuracy of the prediction is correspondingly higher [14].

Finally, traffic accidents are influenced by complex external factors. For example, during peak hours, traffic density plays a crucial role in accident frequency, while during inclement weather conditions, natural environmental factors such as road conditions and obstructed visibility have a more significant impact on crash. Furthermore, compared to the entire urban space, crash data in individual fine-grained regions is quite sparse. This sparse supervisory signal presents a considerable challenge for training deep learning models, as their ability to generate effective spatial–temporal embeddings may be hindered due to the limited data available [15]. Therefore, how to effectively capture and adaptively fuse these multi-source data [16–18], as well as design meaningful self-supervised learning tasks, becomes the key to constructing robust spatial–temporal features for traffic accident data [19].

Motivated by the above observations, we propose an **U**ncertainty-aware spatial–temporal multi-view hypergraph contrastive learning framework for **T**raffic **a**ccident **r**isk prediction (TarU), which combines a deterministic deep learning model with uncertain probabilistic assumptions. To capture the representative features of each region, we use two encoders from local view and global view to learn geographic level and semantic level relationships. Recognizing the importance of leveraging the consistency and complementarity of representations of the same region from different perspectives, we design a multi-view contrastive learning paradigm to facilitate the collaborative training of encoders. This strategy empowers our TarU to construct robust spatial–temporal representations with sparse accident data. In order to explore the influence of different external factors on traffic accidents, a multi-channel attention convolution module is designed to fully extract the correlation between each external information and its influence on traffic accidents. To alleviate spatial imbalance issue, we introduce a hypergraph learning task with learnable hyperedges and hypergraph infomax network to enforce the model discrimination ability. The hyperedges serve as intermediate hubs, linking regions with similar traffic accident patterns for global-aware information passing. Finally, we utilize the zero-inflated negative binomial model to capture the global probabilistic structure of data, addressing the challenges posed by skewed and long-tail distributions.

In summary, the contributions of this work are summarized as follows:

- We propose an integrated framework that combines deterministic deep learning with uncertain probabilistic assumptions for traffic accident risk prediction. Different from existing deep learning methods, we adopt ZINB distribution to solve the zero-inflated problem from a statistical perspective. This framework may shed light on other imbalanced and sparse spatial–temporal prediction problems, such as crime and natural disasters prediction.
- We design a hypergraph learning module that fully addresses the issue caused by imbalanced spatial distribution via learnable hyperedges and hypergraph infomax network.
- We present a multi-view contrastive learning paradigm that allows encoders to collaboratively supervise with each other to promote better integration of complementary features from different perspectives, enforcing the self-discrimination and robustness of TarU in sparse data.
- We conduct extensive experiments on two real-world traffic accident datasets to evaluate our model. The results demonstrate that the proposed TarU significantly outperforms existing methods.

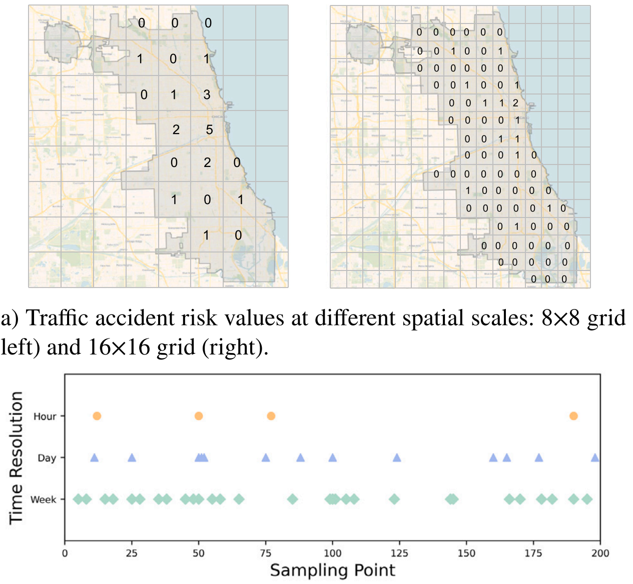
2. Related work

In this section, we will review the relevant work from two aspects: traffic accident risk prediction and methods for unbalanced and sparse data.

2.1. *Traffic accident prediction*

Traffic accident prediction has been an active area of research for several decades and plays a crucial role in various practical applications. In the early stages, traditional statistical methods were widely used, such as Negative Binomial (NB) models and Bayesian networks [20–24]. These methods focused on exploring the relationship between different variables and accident risk, but faced challenges in capturing specific data features, such as the inherent spatial–temporal correlations in traffic data. Later, machine learning models, such as Support Vector Regression (SVR) [25], KNN [26], and decision trees [27], performed well in capturing nonlinear relationships between traffic accidents, leading to better experimental results.

With the emergence of artificial intelligence, deep learning methods have dominated current literature due to their powerful automatic feature extraction and representation capabilities. As a result, researchers have attempted to utilize deep learning-based methods to predict traffic accidents. Chen et al. [28] used Stacked Denoising AutoEncoders (SDAE) to learn hierarchical feature representations of human mobility, exploring the impact of human mobility on traffic accident risk. [29] introduced stacked CNNs to capture spatial correlations between regions, further improving prediction performance. However, both studies overlooked the modeling of temporal correlations in traffic accidents. Therefore, subsequent researchers have employed Recurrent



(a) Traffic accident risk values at different spatial scales: 8×8 grid (left) and 16×16 grid (right).

(b) Traffic accident events at different time resolutions.

Fig. 2. Traffic accident at different spatial scales and time resolutions.

Neural Networks (RNNs) and their variants to capture long-term and short-term dependencies in time [3,4,30]. By analyzing the spatial–temporal patterns of traffic accidents, they built deep models based on recursive neural networks to infer accident risk. [6] used Convolutional Long Short-Term Memory (ConvLSTM) networks to capture the temporal trends and spatial heterogeneity in the data. Given the natural graph structure of traffic networks [31], spatial–temporal GNNs, which combine GNNs with time-series models, have also shown great promise in traffic accident prediction. For instance, [9] constructed a graph based on the fine-grained road structure of the study area and used Graph Convolutional Networks (GCNs) to model spatial information. Liu et al. [5] proposed a traffic accident analysis application framework based on edge computing, using a spatial–temporal graph convolution encoder to capture the dynamic spatial–temporal correlations of traffic conditions and embed them into a low-dimensional vector, followed by a multi-task learning approach to incorporate external factors for traffic accident analysis. In an effort to address the challenges of large-scale road network traffic event prediction, [10] proposed a Multi-structured Graph Neural Network (MSGNN) model, which leverages multiple graphs with different structures to represent various data sources within the same geographic sub-region, thereby facilitating real-time traffic management across the entire network. Despite these advancements, the above-mentioned studies have not adequately addressed the inherent zero-inflation problem in traffic accident data.

Most traffic risk prediction models primarily focus on capturing the spatial–temporal correlations at the geographic level within the data, often overlooking the underlying semantic relationships inherent across the entire urban space. To address this gap, Wang et al. [7] pioneered a traffic accident risk prediction model, GSNet, which incorporates multiple spatial–temporal factors and is capable of modeling the spatial–temporal correlations of traffic accident data from both geographic and semantic perspectives. They proposed three distinct graph structures – risk, road, and points of interest (POI) – to capture the semantic relationships between regions within the graph. Building on this work, Wang et al. [8] introduced a multi-view, multi-task spatial–temporal network model designed for joint prediction of traffic accident risks at both fine-grained and coarse-grained spatial levels. Their approach leverages multi-view self-attention GCNs to facilitate the construction of global semantic features for the nodes in the multi-view spatial graph. However, it is important to note that these studies predominantly focus on static semantic graphs. In contrast, Trirat et al. [32] proposed a method that models both static and dynamic spatial–temporal relationships between regions using heterogeneous features within a multi-view graph framework. They employed an inter-view attention module to adaptively weigh the importance of different semantic features. Unfortunately, these methods all rely on predefined similarity graph structures, which introduce a strong prior bias. In contrast, our proposed hypergraph learning module adaptively learns these features through learnable hyperedges and a hypergraph informax network, which allows for more effective capture of spatial relationships and imbalance issues, thereby improving model robustness.

2.2. Methods for unbalanced and sparse data

The risk of traffic accidents is inherently an imbalanced dataset, typically exhibiting skewed and long-tailed distributions. Additionally, the rarity of traffic accidents results in a large proportion of zero entries in the data. As spatial and temporal resolutions increase, this imbalance and sparsity become even more pronounced. To intuitively demonstrate this phenomenon, we present a visualization of the accident risk values across different spatial resolutions, as shown in Fig. 2(a). It can be observed that as spatial resolution increases (i.e., grids become smaller), the imbalance and sparsity become more prominent. Furthermore, we randomly select a fixed spatial grid and visualize the traffic accident data under different temporal resolutions (e.g., hourly, daily, and weekly), as shown in Fig. 2(b). The results indicate that finer-grained temporal sampling further exacerbates data sparsity, with the proportion of zero entries rising significantly. This zero-inflation characteristic negatively impacts the performance of deep neural networks during training, particularly when non-zero samples in the training labels are extremely rare. In such cases, the model often fails to learn effectively, leading to predictions that are heavily biased towards zero. Previous approaches [5,6,12] have attempted to circumvent this issue by using coarser spatial resolutions or lower temporal resolutions. However, this strategy risks masking critical features of high-risk areas or time periods, thus limiting the ability to gain deeper insights into and accurately predict traffic accident risk.

To address this problem, existing approaches primarily focus on improving model performance through data augmentation and customized loss functions. For example, Zhou et al. [2,33] proposed a prior data augmentation strategy to mitigate the zero-inflation problem. Studies [34,35] adopted Focal loss to address the imbalanced data distribution. However, these approaches mainly focus on anomaly detection tasks (i.e., whether an accident occurs in a given region) and are unable to quantify the severity of accident risk in specific regions.

Recent works [7,36,37] employed weighted loss functions to adjust for samples with higher traffic accident risks. During training, the loss function imposes greater penalties on high-risk samples that are misclassified, thereby increasing the model's sensitivity to these samples and helping to avoid the final prediction being dominated by zero values. Moreover, researchers such as Wang [8] and Trirat [32] proposed using coarse-to-fine constraint losses and Huber loss to alleviate the tendency of deep learning models to predict zeros. However, these methods often implicitly assume homoscedasticity and do not adequately account for the inherent variability of their predictions.

Furthermore, urban events are often spatially sparse, which may lead models to avoid predicting accidents in most areas in favor of achieving lower average errors. This type of prediction is not truly useful for practitioners, such as law enforcement, as the high-risk locations derived from such predictions are likely to be inaccurate. Zhuang et al. [38] were among the first to address the issue of sparse data within the context of uncertainty quantification. They introduced a zero-inflated negative binomial distribution combined with spatial–temporal neural networks to manage the predominant zero instances and non-normal distributions in sparse travel demand data. Their results demonstrated that the proposed model outperforms others when

Table 1
Descriptions of symbols involved in this paper.

Symbol	Description
I, J, R	The row and column dimension in the spatial grid map in a city, number of regions.
T, D	The length of observations, feature dimension of each region at one time step.
r, t	The index of regions, time interval.
$X \in \mathbb{R}^{I\times J\times T\times D}$	Traffic accident features tensor.
Y_t^r	Traffic accident risk at region r, time interval t.
$\sigma, \delta, \varepsilon$	Sigmoid, ReLU, LeakyReLU activation function.
$q_{T+1} \in \mathbb{R}^{d_t}$	The time information of time interval $T+1$, including hour of day, day of week and if it is a holiday.
$Z^{\ell}, Z^{g} \in \mathbb{R}^{R\times d}$	Representations with local and global view encoders.
$E_t \in \mathbb{R}^{R\times d_h}$	Traffic accident embeddings.
$\mathcal{A}_t \in \mathbb{R}^{H\times R}$	Hypergraph dependency structure matrix.
$\Lambda_t \in \mathbb{R}^{R\times d_h}$	Hypergraph-guide region-level representations.
s_t	Graph-level representations.
θ, n, p	The parameters of ZINB distribution.
$\mathcal{L}^I, \mathcal{L}^C, \mathcal{L}^U$	Hypergraph infomax loss, multi-view contrastive loss, negative log likelihood loss.

applied to higher-resolution data. Therefore, it presents a promising solution to combine the zero-inflated distribution and deterministic deep learning technology to address the spatial–temporal uncertainty and sparsity in traffic accident prediction.

In intelligent transportation networks, cooperative perception has emerged as a promising paradigm to overcome the perception limitations of individual intelligent vehicles [39,40]. For example, recent research [41] has tackled the challenge of updating digital twin models in real-time for non-intelligent connected vehicles in mixed traffic environments. Compared to such cooperative perception methods, we focus on data-driven approaches rather than real-time physical sensing. Furthermore, these studies mainly rely on multi-vehicle or multi-sensor information fusion. In contrast, our method adopts a multi-view contrastive learning paradigm, where the contrastive loss enforces collaborative supervision between two encoders. This enables our model to enhance perception capabilities without relying on explicit communication. At the same time, it captures rich information from multiple perspectives, effectively mitigating the issue of data sparsity.

3. Methodology

3.1. Preliminaries

Before introducing our TarU framework, we formulate the traffic accident risk prediction problem. For the convenience of understanding our model, we have listed the related symbols and their descriptions in Table 1.

Geographic Region Feature: We divide the entire urban space into $I \times J$ grids based on longitude and latitude, where each grid represents an equally sized geographical region. At each time step, we collect feature data from each grid that is relevant to the occurrence of traffic accidents, covering multiple factors such as POI, traffic flow, and weather conditions. We use one-hot encoding to represent these features, ultimately generating a feature vector with D dimensions. Then traffic accident feature tensor of all regions at time step t can be represented $X_t \in \mathbb{R}^{I\times J\times D}$.

Traffic Accident Risk: In this study, we define a metric (i.e., traffic accident risk) to evaluate the severity of traffic accidents in a region at a specific time. Based on the number of casualties in crashes, we define three types of traffic accidents, i.e., minor accidents, injury accidents, and fatal accidents, and set the corresponding risk values as 1, 2, and 3, respectively [7]. Y_t^r denotes the risk value of region r at time interval t. For example, region r has two minor accidents, two injury accidents and one fatal accident at time interval t, then $Y_t^r = 2\times1+2\times2+1\times3 = 9$. The traffic accident risk tensor for the entire urban area can be formatted as $Y \in \mathbb{N}^{I\times J\times T}$.

Problem Statement: Given historical observations of traffic accident features $X \in \mathbb{R}^{I\times J\times T\times D}$, our goal is to learn a function f to predict the Probability Mass Function (PMF) of risk value in the next time interval $T+1$, thus analyzing the expectation and confidence intervals of future accident risk. The mapping relationship is expressed as follows:

$$f : (X_1, X_2, \ldots, X_T) \rightarrow PMF(Y_{T+1}). \tag{1}$$

3.2. Framework of TarU

The overview of our proposed model TarU is illustrated in Fig. 3. It consists of two main components: a multi-view spatial–temporal dependency learning module and an uncertainty-aware prediction module, which are responsible for capturing latent embedded representations and decoding probabilistic estimates of future accident risk, respectively.

3.3. Multi-view spatial–temporal dependency learning

First, we model the local spatial–temporal correlations of neighboring regions by using multi-channel attention convolution and a dynamic temporal recurrent module. Then, a hypergraph learning architecture is utilized to adaptively capture higher-order semantic relations based on a parameterized hypergraph structure matrix. To enhance TarU by injecting global information content of the entire city, a hypergraph infomax network is introduced to achieve consistency between node-level and graph-level representations. Finally, we leverage contrastive learning to fulfill self-discrimination among regions, enabling different view encoders to collaboratively supervise and achieve robust spatial–temporal representations of sparse crash data.

3.3.1. Local spatial–temporal view

According to the first law of geography, traffic conditions in geographically close regions are often highly correlated. Therefore, we employ a multi-channel CNN to learn and aggregate the representations of different feature channels on the local receptive field, denoted as:

$$X_t^l = \delta(W_t^l * X_t^{l-1}) + b_t^l, \tag{2}$$

where δ is the ReLU activation function. W^l and b^l are the learnable parameter in the convolution of the lth layer. X_t^l is the output after the convolution of the lth layer at time interval t, and $X_t^0 = X_t$. After the L convolution layers, $X_t^L = \{x_1, x_2, \ldots, x_C\} \in \mathbb{R}^{I\times J\times d_C}$, where d_C represents the compressed channel dimension and each element x_c represents the feature embedding learned on the cth channel. In addition, traffic accident can be influenced by various factors such as road structure design, traffic flow, and meteorological conditions, which can vary across different times and locations. To capture these dynamic feature relationships, we use SENet [42] to adaptively adjust different feature channels. First, we learn the weight coefficients for each channel as follows:

$$\omega = \sigma(W_2^f \delta(W_1^f AvgPool(X_t^L))), \tag{3}$$

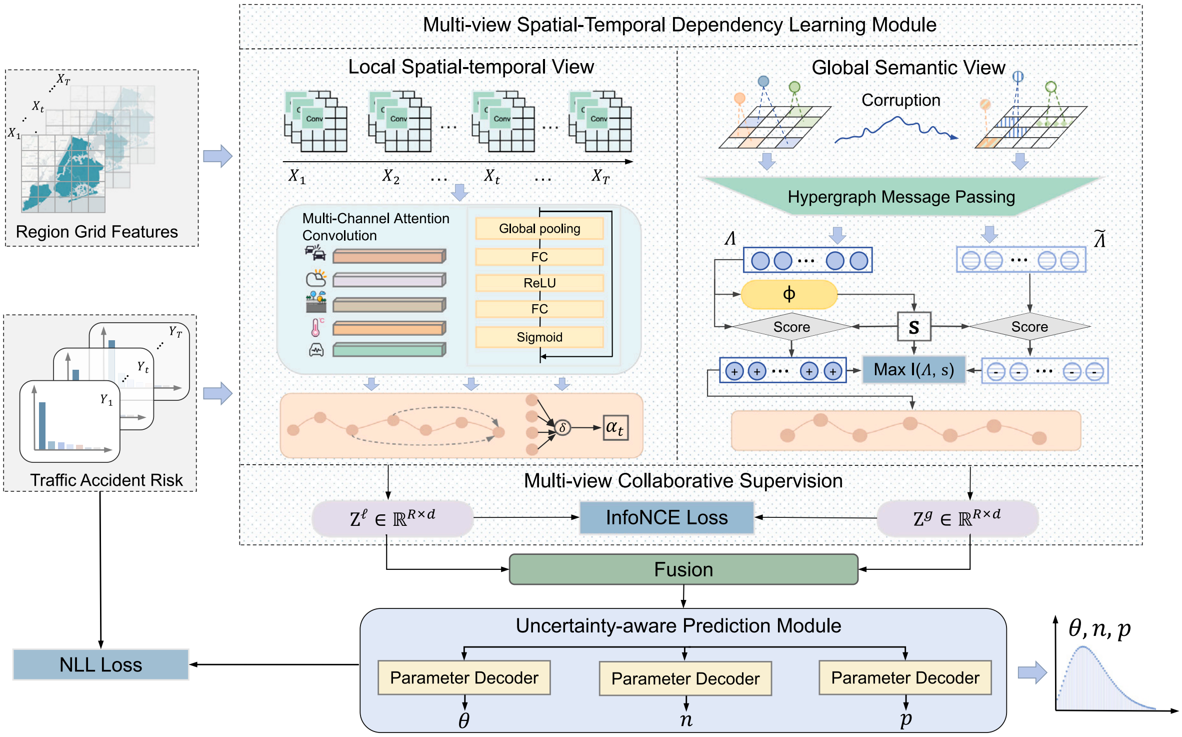


Fig. 3. The framework of TarU.

where $AvgPool(\cdot)$ represents global average pooling. W_1^f and W_2^f are the parameters of the fully-connected layers. σ denotes the sigmoid function. Then, we multiply these coefficients with the original representations to obtain a new weighted representation and add it to the initial input to get the final output $O_t \in \mathbb{R}^{I\times J\times d_C}$.

From a temporal perspective, the inherent periodic nature of traffic data means that accidents are often associated with daily, weekly, and monthly patterns. We extract the time feature from the most recent time ξ slots and the same time interval of the previous η weeks following [7]. Then we use GRU to model both the short-term proximity and long-term periodicity of traffic accidents. The computation process of GRU for region r at time interval t is defined as follows:

$$h_t^r = GRU(O_t^r, h_{t-1}^r), \tag{4}$$

where $h_t^r \in \mathbb{R}^d$ is the hidden state, and d is the number of hidden units. However, the complex temporal autocorrelation varies from case to case. Therefore, we introduce an attention mechanism to dynamically capture temporal patterns:

$$\alpha = softmax(\delta(\mathcal{H}W_H + q_{T+1}W_q + b_\alpha)), \tag{5}$$

where $\mathcal{H} = [h_1, h_2, \ldots, h_\tau]$, $\tau = \xi + \eta$. W_H, W_q and b_α are learnable parameters. δ denotes the ReLU activation function. Given that the occurrence of traffic accidents is influenced by temporal factors such as the hour of the day, the day of the week, and whether it is a holiday, we introduce a time information q_{T+1} for the time step to be predicted. Specifically, this time feature is a 32-dimensional vector: the first 24 dimensions correspond to the 24 h of the day, the next 7 dimensions represent the days of the week (Monday to Sunday), and the final dimension indicates whether the time step falls on a holiday. Finally, we can get the output:

$$Z^\ell = \sum_{i=1}^{\tau} \alpha_i \ h_i, \tag{6}$$

where $Z^\ell \in \mathbb{R}^{R\times d}$, R denotes the total number of grids (i.e. $R = I\times J$). In this way, local geographic information and dynamic time dependencies are integrated in Z^ℓ.

3.3.2. Global semantic view

In addition to the influence of adjacent areas, semantic dependence implied in the entire urban space is also crucial for accurately forecasting traffic accident risk. For example, intersections in two business districts, though geographically distant, may exhibit similar traffic accident patterns due to their comparable urban functional attributes and road design. Therefore, we design a trainable hypergraph structure matrix to automatically derive cross-region semantic dependencies at the global level. Meanwhile, we introduce a hypergraph infomax network with mutual information consistency between node-level and graph-level spatial–temporal representations. Hypergraph learning task enhances TarU to address the challenge of imbalanced accident distributions under sparse conditions.

A hypergraph consists of a set of vertices and hyperedges, where each hyperedge can connect an arbitrary number of vertices, which shows a unique advantage over pairwise relations with uniform nodes [43]. Therefore, we empower TarU to capture complex higher-order relationships among regions with a hypergraph message-passing mechanism [12,44]:

$$\Lambda_t = \varepsilon(\mathcal{A}_t^\top \cdot \varepsilon(\mathcal{A}_t \cdot E_t)), \tag{7}$$

where $E_t \in \mathbb{R}^{R\times d_h}$ is the embedding for traffic accident risk obtained by augmenting the original data X_t with a linear layer and d_h is the embedding dimension. $\mathcal{A}_t \in \mathbb{R}^{H\times R}$ denotes the learnable hypergraph dependency structure matrix and H is the number of hyperedges. ε is the LeakyReLU activation function. It can be seen that $\mathcal{A}_t$ is related to the tth time slot, allowing our model to capture dynamic features in the global evolution. Furthermore, $\mathcal{A}$ preserves the global urban context, enabling regions with similar accident patterns to be correlated, thereby mitigating spatial distribution imbalance issue.

To incorporate global relationships into the representation of each region for enhancing the self-supervised signal, we introduce a hypergraph infomax network inspired by [45]. First, we utilize a readout function to obtain a graph-level summary representation:

$$s_t = \phi(\Lambda_t) = \sum_{r=1}^{R} \Lambda_t^r / R, \tag{8}$$

where $s_t \in \mathbb{R}^{d_h}$ aggregates the global information of the whole graph at time interval t. Then, we employ a corruption function to obtain region-level representations from corrupted hypergraph structures as negative samples:

$$\tilde{\Lambda}_t = \mathrm{Corruption}(\varepsilon(\tilde{\mathcal{A}}_t^\top \cdot \varepsilon(\tilde{\mathcal{A}}_t \cdot \tilde{\mathcal{E}}_t))). \tag{9}$$

Here we employ randomly disrupt the region indexes for corruption. Next, we use a score function to evaluate the region-level representation to determine whether the node is from the original hypergraph or the corrupt hypergraph:

$$\gamma_t^r = \mathrm{Score}(\Lambda_t^r, s_t) = \sigma((\Lambda_t^r)^\top W^b s_t), \tag{10}$$

$$\tilde{\gamma}_t^r = \mathrm{Score}(\tilde{\Lambda}_t^r, s_t) = \sigma((\tilde{\Lambda}_t^r)^\top W^b s_t). \tag{11}$$

We use the bilinear scoring function, where $W^b \in \mathbb{R}^{d_h \times d_h}$ is a learnable scoring matrix and σ is the sigmoid function used to convert the scores to probabilities of (Λ_t^r, s_t) being a positive sample. To seek a node representation that captures global information about the entire hypergraph, we use a noise-contrastive type objective:

$$\mathcal{L}^I = -\sum_{r=1}^{R} \left(\log \gamma_t^r + \log(1 - \tilde{\gamma}_t^r)\right). \tag{12}$$

Finally, the output of the global semantic view, $\mathrm{Z}^g \in \mathbb{R}^{R \times d}$, is obtained through the same temporal recurrent module used for the local spatial–temporal view.

3.3.3. Multi-view collaborative supervision training

In the latent embedding space, contrastive learning is introduced to learn the consistent representations across different views of the same region while distinguishing the noise involved in different regions [46]. In the context of traffic accident risk prediction, for the same region, features learned from the local perspective should be as similar as possible to the features learned from the global perspective, as they describe the same region. However, for different regions, their features should be as far apart as possible. Therefore, a contrastive loss function can be employed to strengthen the consistency of features within the same region and the disparity of features across different regions. We treat the same region from different views as positive pairs and different regions as negative pairs. The contrastive loss at time interval t is defined as follows:

$$\mathcal{L}^C = -\sum_{r=1}^{R} \Bigg[\log\left(\exp\left(\cos(\mathrm{Z}_{t,r}^{\ell}, \mathrm{Z}_{t,r}^{g})/\Delta\right)\right) - \log\left(\sum_{r'=1}^{R} \exp\left(\cos(\mathrm{Z}_{t,r}^{\ell}, \mathrm{Z}_{t,r'}^{g})/\Delta\right)\right)\Bigg], \tag{13}$$

where $\cos(\cdot)$ is the cosine similarity function. The temperature parameter Δ controls the similarity difference between positive and negative samples. With the regularization of $\mathcal{L}^C$, our model enhances the self-discrimination of accident occurrence patterns across different regions and time periods, thereby achieving robust representations of traffic accident data for sparse interactions. Finally, we use a fully-connected layer to fuse the spatial–temporal embeddings obtained from the two views.

3.4. Uncertainty-aware prediction

The data distribution of traffic accident risk generally conforms to three characteristics: (1) discrete; (2) variance greater than the mean; (3) rich in zeros, which deviates from the Gaussian assumption underlying previous deterministic deep learning models [38,47]. In this case, distribution such as the ZINB is more appropriate. Compared to the Poisson distribution, the ZINB model introduces a Negative Binomial (NB) distribution to model the occurrence of events, thereby more effectively handling the overdispersion and uncertainty in the data. Therefore, to better capture the zero-inflation and long-tail non-zero characteristics of data, we integrate the ZINB model into TarU and decode the latent embedded representation Z to the probabilistic estimation of future accident.

3.4.1. Zero-inflated negative binomial distribution

The process of generating data with the ZINB distribution involves two steps: (1) they are either zeros with probability θ, (2) or non-zero values with probability $1 - \theta$ following the NB distribution. Therefore, the PMF of ZINB is denoted as:

$$f_{ZINB}(y|\theta, n, p) = \begin{cases} \theta + (1-\theta) f_{NB}(0|n, p) & \text{if } y = 0 \\ (1-\theta) f_{NB}(y|n, p) & \text{if } y > 0, \end{cases} \tag{14}$$

$$f_{NB}(y|n, p) = \binom{y+n-1}{n-1} (1-p)^y p^n, \tag{15}$$

where sparsity parameter θ is utilized to denote the proportion of zero values, n and p are the shape parameters of the NB distribution that determine the number of successes and the probability of a single failure respectively.

3.4.2. Parameter decoder

We design three parameter decoders to convert the latent representations of traffic accidents into the parameters of ZINB distribution:

$$P(*) = act(W_* Z + b_*). \tag{16}$$

We use the softplus activation function to ensure that the parameter n remains non-negative. For the parameters θ and p, we employ the sigmoid function to constrain their values within the range $(0, 1)$, representing probabilities. Here, W_* and b_* denote trainable parameters. The learning objective of the ZINB model is to maximize the log likelihood function:

$$LL_y = \begin{cases} \log\theta + \log(1-\theta) p^n & \text{if } y = 0 \\ \\ \log(1-\theta) + \log\Gamma(n+y) & \\ -\log\Gamma(y+1) - \log\Gamma(n) & \\ +n\log p + y\log(1-p) & \text{if } y > 0 \end{cases}, \tag{17}$$

where y is the ground truth and Γ denotes the Gamma function. Finally, we use the negative log likelihood as our uncertainty-aware loss function:

$$\mathcal{L}^U = -LL_{y=0} - LL_{y>0}. \tag{18}$$

By leveraging the flexibility of uncertainty-aware prediction module, our model can be extended to accommodate other distributions by using Z to represent the corresponding shape parameter sets.

3.5. Model training

The final objective function is defined as:

$$\mathcal{L} = \mathcal{L}^{\mathrm{I}} + \lambda_c \mathcal{L}^{\mathrm{C}} + \lambda_u \mathcal{L}^{\mathrm{U}}, \tag{19}$$

where λ_c and λ_u are parameters controlling the weights to balance the losses. Algorithm 1 presents the pseudocode for the main training process of TarU, detailing the key steps and procedures involved in its implementation.

4. Experiments

In this section, we presents the evaluation and analysis of TarU. We conduct a series of experiments to assess the performance of TarU and answer the following four questions:

- Q1: **Superiority**. Does TarU outperform current state-of-the-art traffic accident prediction models?
- Q2: **Effectiveness**. Are the proposed modules effective in mitigating data imbalance and sparsity, ensuring robust representation?
- Q3: **Sensitivity**. How sensitive is TarU's performance to variations in hyperparameters?
- Q4: **Interpretability**. Can the visual outputs of TarU enhance the interpretability of model predictions?

Algorithm 1: Training Algorithm of TarU

```
Input: Traffic accident features $X \in \mathbb{R}^{I\times J\times T\times D}$, time
       information $q_{T+1}$, maximum epoch number $E$, loss
       weight $\lambda_c$, $\lambda_u$, learning rate $lr$
Initialize model parameters $\Psi$.
for $i = 1$ to $E$ do
    Compute $O_t$ using multi-channel attention convolutions
      according to Eq. 1-3.
    Encode $h$ via GRU according to Eq. 4.
    Compute attention coefficient $\alpha$ according to Eq. 5.
    Compute local spatial-temporal view output $\mathrm{Z}^{\ell}$ according
      to Eq. 6.
    Encode global higher-order relationships $\Lambda$ via hypergraph
      neural network according to Eq. 7.
    Compute graph-level summary representation $s$ via readout
      function according to Eq. 8.
    Generate region-level representations $\tilde{\Lambda}$ from corrupted
      hypergraph structures according to Eq. 9.
    Compute node scores for original and corrupted
      hypergraphs according to Eq. 10-11.
    Compute global semantic view output $\mathrm{Z}^{g}$.
    Fuse $\mathrm{Z}^{\ell}$ and $\mathrm{Z}^{g}$ via fully-connected layer.
    Compute ZINB distribution parameters $\theta, n, p$ according to
      Eq. 16.
    Compute infomax loss $\mathcal{L}^{\mathrm{I}}$ according to Eq. 12.
    Compute contrastive loss $\mathcal{L}^{\mathrm{C}}$ according to Eq. 13.
    Compute uncertainty-aware loss $\mathcal{L}^{\mathrm{U}}$ according to Eq. 17-18.
    Compute total loss $\mathcal{L}$ according to Eq. 19.
    for $\psi \in \Psi$ do
        $\psi = \psi - lr \cdot \partial\mathcal{L}/\partial\psi$
    end for
end for
Output: The final trained model parameters $\Psi$
```

- Q5: **Efficiency**. How efficient is the TarU framework compared to other baselines?

In the following subsection, we first present the experimental setup and then report the evaluation results addressing the research questions outlined above.

4.1. Experimental settings

4.1.1. Datasets

This study utilizes two real-world traffic accident datasets from NYC[2] and Chicago[3] to validate the effectiveness of our model. To preprocess the multi-source data for traffic accident risk prediction, we first partition the urban areas of New York City and Chicago into 2 km × 2 km grid cells based on geographical coordinates (longitude and latitude). Since certain areas lack road infrastructure, we retain 243 grid cells in New York City and 197 in Chicago that contain road networks to ensure meaningful risk prediction. For the traffic accident risk data, we apply the risk calculation method described in Section 3.1. The temporal features used in the model include the hour of the accident, the day of the week, and a holiday indicator, with all time-related variables represented using one-hot encoding. The weather data includes temperature and weather conditions. Temperature is treated as a continuous variable, while weather conditions (sunny, rainy, snowy, cloudy, and foggy) are encoded as one-hot vectors. For POI data, we categorize locations into seven types, including residential areas, schools, cultural facilities, entertainment venues, social services, transportation hubs, and commercial areas. The number of POIs for each category is counted within each grid cell and encoded as a 7-dimensional vector. The taxi trip data is processed by matching pick-up and drop-off locations to the corresponding grid cells and time steps. All datasets are processed with an hourly temporal resolution. To ensure chronological integrity, we split the data into training, validation, and test sets in a 6:2:2 ratio, following a strictly time-ordered manner. Detailed statistics of the dataset are given in Table 2.

Table 2
Dataset description.

Dataset	NYC	Chicago
Time span	Jan, 2013 to Dec, 2013	Feb, 2016 to Sep, 2016
Accidents	147 k	44 k
Taxi trips	173,179 k	1744 k
POI data	15,625	–
Weather	8760	5,832
Spatial unit	2 km ×2 km	2 km ×2 km
Grid scale	20 × 20	20 × 20
Zero rate	93.73%	97.81%

4.1.2. Implementation details

We implemented TarU in the Pytorch framework and used the Adam optimizer for training. During hyperparameter tuning, we employ a grid search method to explore the optimal combination of parameters and select the best configuration based on its performance on the validation set. In multi-channel attention convolutional network, the convolutional kernel size is set to 3 and two convolutional layers are used. For the temporal recurrent module, the length of adjacent time interval $\xi = 4$ and weeks $\eta = 3$, and the hidden state $d = 256$ in GRU. In the hypergraph learning module, the dimension of embedding $d_h = 16$, the number of hyperedges $H = 128$. We set the hyperparameters λ_c and λ_u in the final objective loss at 1. We conduct our experiments on a machine with NVIDIA GeForce RTX 3090 GPU.

4.1.3. Metrics

We use Recall and MAP following [7,48] to evaluate the accuracy of predicting high-risk accident regions. For time interval t, if there are $|R_t|$ regions actually occur accident, then Recall indicates the percentage of the top $|R_t|$ predicted high-risk areas that overlap with the actual accident locations. The specific formula is as follows:

$$\mathrm{Recall} = \frac{1}{T}\sum_{t=1}^{T}\frac{P_t \cap R_t}{|R_t|}, \tag{20}$$

where R_t is the set of regions where traffic accidents actually occurred at time interval t and $|R_t|$ represents the number of regions. P_t denotes the set of the top $|R_t|$ high-risk regions predicted at time t. Recall measures the percentage of model predictions that hit in areas where accidents actually occur, so a higher score represents a better identification of high-risk regions. MAP is used to evaluate how closely the predicted ranking of high-risk regions aligns with the ranking of actual accident regions. The formula is as follows:

$$\mathrm{MAP} = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{j=1}^{|R_t|} pre(j)\times rel(j)}{|R_t|}, \tag{21}$$

where $pre(j)$ denotes the precision of a cut-off rank list from 1 to j, and $rel(j)$ is the recall of region j. If there are traffic accidents in the region j, then $rel(j) = 1$, otherwise $rel(j) = 0$. MAP represents the average precision and higher MAP denotes better performance of the model.

To assess the uncertainty of the outcomes, we use the Prediction Interval Coverage Probability (PICP) and Mean Prediction Interval Width (MPIW) introduced in previous work [38,49] on the 10%–90%

[2] https://opendata.cityofnewyork.us/

[3] https://data.cityofchicago.org/

Table 3
Performance comparisons on two datasets. The best results are marked in bold and the suboptimal results are marked by the asterisk. '-' means the model output is not available.

Model	NYC				Chicago			
	Recall	MAP	PICP	MPIW	Recall	MAP	PICP	MPIW
SVM(2011)	24.80%	0.0821	–	–	12.29%	0.0416	–	–
GRU(2014)	30.11%	0.1553	–	–	18.17%	0.0698	–	–
ConvLSTM(2015)	31.98%	0.1638	–	–	19.27%	0.0753	–	–
Hetero-ConvLSTM(2018)	31.54%	0.1584	–	–	18.82%	0.0732	–	–
GSNet(2021)	33.16%	0.1787	–	–	19.92%	0.0877	–	–
ST-HSL(2022)	33.95%	0.1874	–	–	20.92%	0.0901	–	–
MVMT-STN(2023)	33.78%	0.1887	–	–	20.72%	0.0928	–	–
C-ViT(2023)	33.86%	0.1875	–	–	20.93%	0.0980	–	–
TWCCnet(2024)	33.48%	0.1864	–	–	20.21%	0.0845	–	–
AGSSL(2024)	32.26%	0.1676	–	–	18.82%	0.0734	–	–
MGHSTN(2024)	34.50%	0.1903	–	–	21.21%*	0.1004*	–	–
TarU-Gaussian(ours)	33.16%	0.1791	0.6026	0.2901	19.76%	0.0915	0.9348	0.3241
TarU-NB(ours)	34.55%	0.1898	0.9621	0.0288	20.87%	0.0994	0.9546	0.0471
TarU-ZIP(ours)	34.70%*	0.1904*	0.9713*	0.0138*	20.97%	0.0998	0.9589*	0.0131*
TarU(ours)	**35.21%**	**0.1980**	**0.9813**	**0.0102**	**22.46%**	**0.1128**	**0.9716**	**0.0053**

confidence interval. PICP assesses whether the prediction intervals accurately capture the ground truth:

$$\text{PICP} = \frac{C_{obj}}{T}, C_{obj} = \sum_{t=1}^{T} I\ \{\hat{L}_t < Y_t < \hat{U}_t\}. \tag{22}$$

where $I\{\}$ is an indicator function. Y_t is the true value of all regions at time interval t. L and U represent the upper and lower bounds, respectively. However, it does not make sense to improve the PICP simply by increasing the interval width. Therefore, we further introduce MPIW:

$$\text{MPIW} = \frac{1}{T}\sum_{t=1}^{T}(\hat{U}_t - \hat{L}_t). \tag{23}$$

Larger coverage probability and smaller prediction interval are more desirable.

4.1.4. Baselines

In order to explore the advantages of TarU, we compared it with the following representative traffic accident risk prediction methods.

(1) SVM [50]: SVM predicts traffic accident data using support vector machines.
(2) GRU [51]: GRU can effectively deal with long-term dependencies in time-series data.
(3) ConvLSTM [52]: ConvLSTM is a model that combines CNN and LSTM to capture the temporal autocorrelation and the local spatial features of accidents.
(4) Hetero-ConvLSTM [6]: Hetero-ConvLSTM incorporates spatial graph features and spatial model ensemble on the basis of ConvLSTM to effectively capture temporal trends and spatial heterogeneity in data.
(5) GSNet [7]: GSNet learns the spatial–temporal correlation from both geographic and semantic aspects. The algorithm employs CNN and GCN to capture neighboring geographic regions and global semantic correlations. They design weight loss to solve the zero-inflation problem.
(6) ST-HSL [12]: ST-HSL learns spatial–temporal representation by unifying hypergraph dependency modeling with self-supervision learning.
(7) MVMT-STN [8]: A multi-view, multi-task spatial–temporal network model designed for joint prediction of traffic accident risks at both fine-grained and coarse-grained spatial levels.
(8) C-ViT [36]: This model utilizes vision transformer and reformulates the traffic accident risk prediction problem as image regression problem.
(9) TWCCnet [53]: This framework uses department semantic associations to dynamically weight multiple contextual factors, providing spatial–temporal correlations of traffic accidents from both neighborhood and semantic perspectives.
(10) AGSSL [35]: This method combines adaptive graph and self-supervised learning to predict traffic accidents, captures spatial correlation of urban areas through adaptive graph structure, and uses graph information maximization and focus contrast regularization to deal with data imbalance.
(11) MGHSTN [54]: MGHSTN introduces remote sensing images to capture regional backgrounds and employs a multi-granularity hierarchical spatial–temporal network to predict accident risk.

Moreover, we compare our approach with other probabilistic assumptions. Specially, we construct TarU-Gaussian, TarU-NB, TarU-ZIP by replacing the probability assumptions in the uncertainty-aware prediction module with Gaussian, NB distributions and Zero-Inflated Poisson (ZIP) distribution, respectively. These three distributions have only two parameters, while TarU has three. The other components and parameters are the same in these models.

4.2. Performance comparison (RQ1)

Table 3 presents the prediction performance of various methods on the NYC and Chicago datasets. It is clear that our TarU model consistently outperforms all other methods across all evaluation metrics for both datasets. Notably, the performance improvement of our model is more pronounced on the Chicago dataset, where we observe a 5.89% improvement in Recall and a 12.35% improvement in MAP. This substantial gain can be attributed to the higher sparsity of the Chicago dataset, which contains approximately 97.81% zero-value records, as shown in Table 2. This result further validates the effectiveness of TarU's hypergraph-enhanced contrastive learning and cross-view collaborative supervision.

Traditional SVM methods have limitations in modeling complex dependencies in spatial–temporal data, while deep learning-based GRU models have achieved better performance in this task. However, none of these methods have been able to effectively and simultaneously model spatial–temporal correlations in traffic accident data. Although ConvLSTM and HeteroConvLSTM are able to take into account both temporal and spatial correlations, they still neglect the dual imbalance problem in the data. GSNet, MVMT-STN, C-ViT, and TWCCNet address the category imbalance problem by customizing the loss function, but these methods fail to adequately take into account the spatial distribution of accidents imbalance and complex dynamic processes. However, our proposed TarU model utilizes a learnable hypergraph dependency

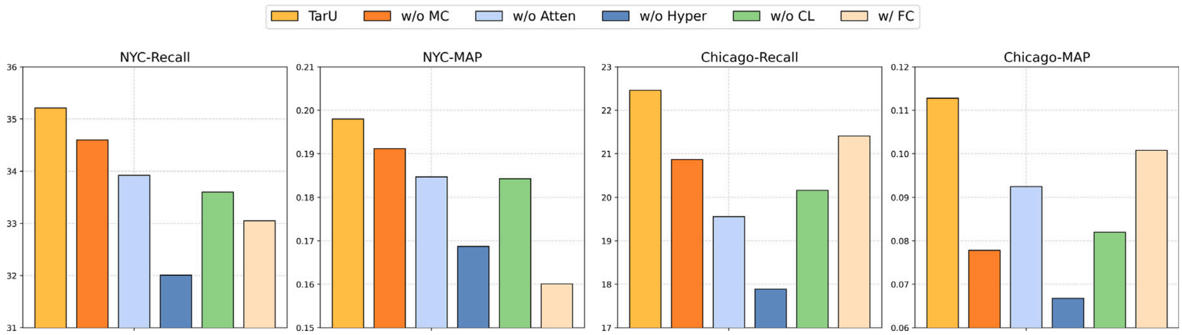


Fig. 4. Ablation study on two datasets.

structure matrix, which can capture the dynamic features of the global evolutionary process more efficiently.

AGSSL was originally designed for the anomaly detection task as a binary classification problem (i.e., predicting whether or not an accident has occurred), and thus it is out of its league for our accident risk prediction task. ST-HSL employs a hypergraph structure to capture dynamic high-order dependencies and MGHSTN utilizes remotely sensed imagery as auxiliary information to enhance traffic accident characterization. However, the sporadic nature of traffic accident data and the extremely sparse supervisory information inevitably cause these models to generate suboptimal spatial–temporal relational embeddings. In contrast, TarU leverages its multi-channel attention convolution, hypergraph augmentation network, and contrastive self-supervised optimization strategies, allowing the model to more effectively capture complex dependencies and handle the sparse supervised signals in traffic accident data.

For discrete and sparse traffic risk data, TarU-Gaussian fails to capture the skewed data distribution, leading to lower Recall and MAP. In contrast, the negative binomial distribution in the probabilistic layer effectively captures the discrete nature of the data, making the TarU-NB and TarU models more capable of accurately predicting traffic accident risk. Additionally, both TarU and TarU-ZIP successfully learn the sparsity of the data, resulting in very small MPIW values. The ZIP model assumes that there are too many zeros in the data and treats them as an independent process. The ZINB model, on the other hand, further incorporates the overdispersion of the negative binomial distribution, allowing TarU to better handle the volatility and heterogeneity in traffic accident risk data.

The experimental results demonstrate that our model outperforms existing methods in prediction accuracy, enabling more precise identification of high-risk areas. This advantage can be directly applied to optimize police force deployment and traffic signal control, assisting traffic management departments in implementing targeted interventions to reduce traffic accident rates. Furthermore, our model is able to quantify the uncertainty of the prediction, and the system can adopt more conservative strategies in high uncertainty regions, such as increasing the frequency of warnings or strengthening monitoring, rather than relying solely on fixed thresholds to trigger interventions.

4.3. Ablation study (RQ2)

We further analyze the effectiveness of various components of the TarU. Specifically, we design five different variants: (1) w/o MC: We disable the multi-channel attention convolution module. (2) w/o Atten: In this variant, we remove the attention mechanism of time recurrent module. (3) w/o Hyper: We do not explore the hypergraph relationships of the regions, relying solely on local spatial relationship modeling for prediction. (4) w/o CL: We do not use multi-view contrastive learning paradigm for collaborative supervision training. (5) w/FC: We use a fully-connected layer to decode the spatial–temporal embeddings without ZINB-based assumption. The outcomes from Fig. 4 reveal:

- The w/o MC variant exhibits worse performance than TarU, demonstrating the critical role of our multi-channel attention convolution module in discovering dynamic correlations in the causes of traffic accidents.
- The poorer performance of w/o Atten compared to TarU highlights that dynamic temporal dependencies are crucial for predicting traffic accidents.
- The w/o Hyper variant achieves the lowest performance in most cases. This can be explained from two perspectives: (a) The learnable hypergraph structure can adaptively capture global semantic information across the entire city space. (b) The hypergraph infomax network is designed to maximize the mutual information between all node-level representations and the global graph summary, which mitigates the challenges posed by imbalanced spatial data distribution and sparse supervisory information.
- The w/o CL variant performs worse than TarU, supporting the notion that multi-view collaborative supervision training helps mitigate the sparsity issue and enhance the latent representation of traffic accident data.
- The results of w/FC further demonstrate that our model can better fit the discrete distribution of traffic accident risk data.

4.4. Parameter analysis (RQ3)

The parameters λ_c and λ_u determine the weights of various losses. We vary their values in different ranges from {0.001, 0.01, 0.1, 1, 10} to evaluate the impact of λ_c and λ_u on model performance. As shown in Fig. 5, TarU achieves satisfactory results when both λ_c and λ_u are set to 1. Furthermore, we find that when more data is available (e.g., in the NYC dataset), the model becomes less sensitive to hyperparameter variations. In contrast, for smaller and sparser datasets (e.g., the Chicago dataset), the model exhibits higher sensitivity to hyperparameter choices, suggesting that careful tuning is necessary to achieve optimal performance in such scenarios.

We conduct experiments on embedding dimension within the range of {4, 8, 16, 32, 64}. According to Fig. 6(a), the embedding dimension $d_h = 16$ achieves the best performance. Larger dimensions may lead to overfitting in the spatial–temporal representation of the accident data. We also search within the range of {32, 64, 128, 256, 512} for the number of hyperedges H. The experimental results shown in Fig. 6(b) demonstrate that when $H = 128$, the hypergraph effectively captures the global cross-regional dependencies of accident occurrence patterns.

4.5. Case study (RQ4)

To further assess our model's ability to predict traffic accident risk, we visualize the predicted outcomes and the ground truth, as shown in Fig. 7. It can be seen that accidents are truly sparse and there is a very small percentage of high-risk regions. Furthermore, the heat maps of the predicted results closely match the ground truth across both datasets, further validating the effectiveness of TarU.

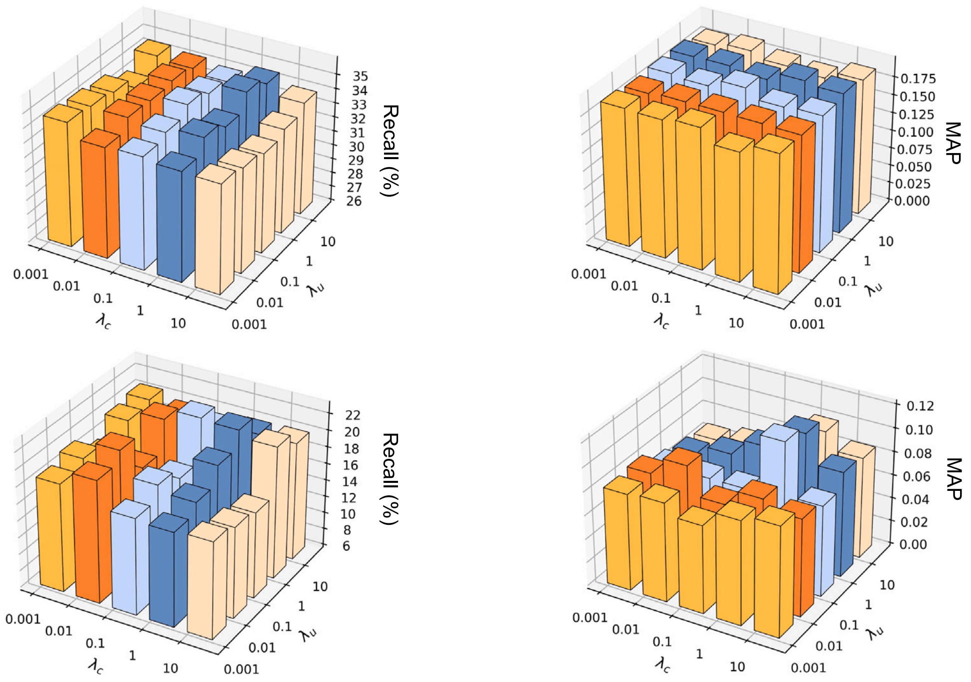
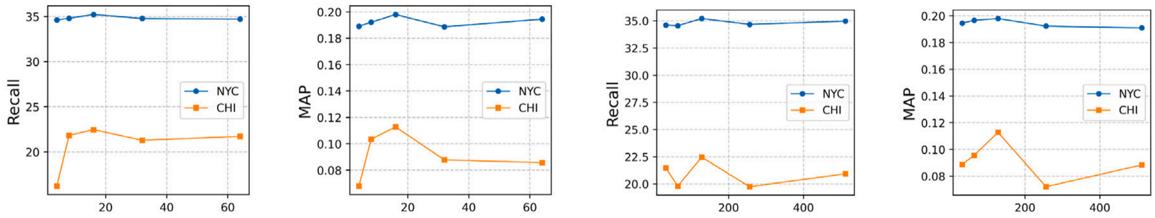


Fig. 5. Parameter analysis on both datasets. The two sub-figures above show the experimental results on the NYC dataset, corresponding to the impact analysis of λ_c and λ_u respectively; the two sub-figures below are the results on the Chicago dataset.



(a) Number of Embedding Dimension d_h.

(b) Number of Hyperedge H.

Fig. 6. Analysis of hyperparameters in terms of embedding dimension and hyperedge number.

Table 4
Efficiency analysis of different models.

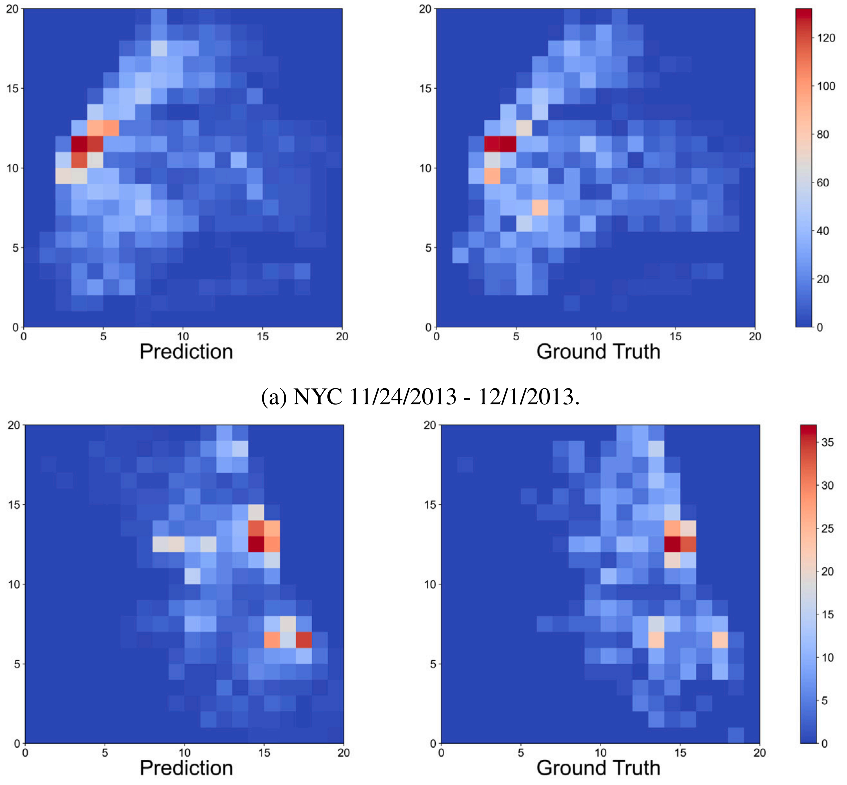
Model	Parameter (M)		GPU memory (GB)		Training time (s/epoch)		Inference time (s)	
	NYC	Chicago	NYC	Chicago	NYC	Chicago	NYC	Chicago
GSNet	6.28	3.76	6.93	5.59	12.82	6.33	1.52	0.82
ST-HSL	3.09	3.09	3.37	3.37	10.62	5.74	3.75	3.51
MVMT-STN	5.95	3.13	7.82	6.46	12.79	5.56	1.93	1.08
C-ViT	0.29	0.29	1.22	1.22	5.12	4.57	0.89	0.55
TWCCnet	3.95	3.46	11.99	5.05	17.47	8.53	2.49	1.41s
MGHSTN	3.98	4.76	8.58	9.97	27.44	22.79	2.76	1.33
TarU	1.65	1.62	5.48	5.48	8.17	4.97	1.28	0.71

4.6. Model efficiency study (RQ5)

We evaluate the scalability of the TarU framework in comparison to state-of-the-art techniques. All experiments are conducted using the default parameter settings on a machine. To assess the efficiency of the methods, we compare key metrics including the number of parameters, memory usage, training time per epoch, and inference time on both the NYC and Chicago datasets. The experimental results are presented in Table 4. We can see that TarU outperforms most baseline methods, second only to C-ViT. While C-ViT gains efficiency by reformulating the problem as an image regression task, TarU shows a clear advantage in prediction accuracy. On the NYC dataset, TarU improves by 4% and 5.6%, while on the Chicago dataset, the improvements are 7.31% and 15.1%, respectively. Additionally, TarU's inference time is nearly halved compared to the best-performing baseline, MGHSTN, highlighting its efficiency and potential for real-time scenarios. For example, our framework can provide real-time traffic accident prediction information to autonomous driving systems [55], enabling self-driving vehicles to optimize their driving strategies in real-time, proactively avoid potential danger zones, and enhance overall safety.

5. Conclusion

In this paper, we propose a novel TarU framework that combines deterministic deep learning methods with uncertain probabilistic assumptions to forecast traffic accident risk. To address class imbalance



(a) NYC 11/24/2013 - 12/1/2013.

(b) Chicago 9/11/2016 - 9/18/2016.

Fig. 7. Comparison of predicted and ground truth traffic accident risk in NYC and Chicago.

issue caused by rich zeros, we introduce ZINB with a sparsity parameter θ to learn the likelihood of the inputs being zero. To tackle spatial distribution imbalance, we design a hypergraph learning task that employs hyperedge-based embedding propagation and hypergraph infomax network to capture non-pairwise dynamic interactions and strengthen region representations. Furthermore, we develop a multi-view contrastive learning paradigm to improve the discriminative power and robustness of TarU, effectively mitigating the challenge caused by inherent rareness of accidents. Extensive experiments on two real-world datasets demonstrate the superiority of TarU over competitive baseline models.

Our work not only provides a new approach to traffic accident prediction from a theoretical perspective but also has the potential to extend to other prediction tasks involving imbalanced and extremely sparse data, such as high spatial–temporal resolution natural disaster forecasting and crime prediction. In practical applications, our framework demonstrates tremendous potential. By deeply integrating with digital twins and intelligent transportation systems [55], and combining real-time traffic flow and weather forecast data, our framework enables real-time traffic accident prediction and dynamic early warning, optimizing traffic flow and resource scheduling. Moreover, we believe that integrating more diverse data sources (e.g., real-time traffic camera feeds, driver behavior data) and exploring more advanced uncertainty modeling techniques will further enhance the robustness and reliability of traffic accident risk prediction.

CRediT authorship contribution statement

Yimei Zhang: Writing – review & editing, Writing – original draft, Methodology. **Guojiang Shen:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition. **Wenyi Zhang:** Writing – review & editing, Visualization, Validation. **Kaili Ning:** Writing – review & editing, Visualization, Validation, Data curation. **Renhe Jiang:** Visualization, Validation, Formal analysis. **Xiangjie Kong:** Writing – review & editing, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62476247, 62073295 and 62072409, in part by the “Pioneer” and “Leading Goose” R&D Program of Zhejiang under Grant 2024C01214, and in part by the Zhejiang Provincial Natural Science Foundation, China under Grant LR21F020003.

Data availability

The authors do not have permission to share data.

References

[1] X. Kong, J. Wang, Z. Hu, Y. He, X. Zhao, G. Shen, Mobile trajectory anomaly detection: Taxonomy, methodology, challenges, and directions, IEEE Internet Things J. 11 (11) (2024) 19210–19231.

[2] Z. Zhou, Y. Wang, X. Xie, L. Chen, H. Liu, RiskOracle: A minute-level citywide traffic accident forecasting framework, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, 2020, pp. 1258–1265.

[3] J. Bao, P. Liu, S.V. Ukkusuri, A spatiotemporal deep learning approach for citywide short-term crash risk prediction with multi-source data, Accid. Anal. Prev. 122 (2019) 239–254.

[4] S. Moosavi, M.H. Samavatian, S. Parthasarathy, R. Teodorescu, R. Ramnath, Accident risk prediction based on heterogeneous sparse data: New dataset and insights, in: Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, 2019, pp. 33–42.

[5] Z. Liu, Y. Chen, F. Xia, J. Bian, B. Zhu, G. Shen, X. Kong, TAP: Traffic accident profiling via multi-task spatio-temporal graph representation learning, ACM Trans. Knowl. Discov. Data 17 (4) (2023).

[6] Z. Yuan, X. Zhou, T. Yang, Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2018, pp. 984–992.

[7] B. Wang, Y. Lin, S. Guo, H. Wan, GSNet: Learning spatial-temporal correlations from geographical and semantic aspects for traffic accident risk forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, 2021, pp. 4402–4409.

[8] S. Wang, J. Zhang, J. Li, H. Miao, J. Cao, Traffic accident risk prediction via multi-view multi-task spatio-temporal networks, IEEE Trans. Knowl. Data Eng. 35 (12) (2023) 12323–12336.

[9] L. Yu, B. Du, X. Hu, L. Sun, L. Han, W. Lv, Deep spatio-temporal graph convolutional network for traffic accident prediction, Neurocomputing 423 (2021) 135–147.

[10] T. Tran, D. He, J. Kim, M. Hickman, MSGNN: A multi-structured graph neural network model for real-time incident prediction in large traffic networks, Transp. Res. Part C Emerg. Technol. 156 (2023) 104354.

[11] B. An, X. Zhou, Y. Zhong, T. Yang, SpatialRank: Urban event ranking with NDCG optimization on spatiotemporal data, in: Advances in Neural Information Processing Systems, vol. 36, 2024, pp. 9919–9930.

[12] Z. Li, C. Huang, L. Xia, Y. Xu, J. Pei, Spatial-temporal hypergraph self-supervised learning for crime prediction, in: 2022 IEEE 38th International Conference on Data Engineering, ICDE, 2022, pp. 2984–2996.

[13] Y. Zhang, X. Kong, W. Zhou, J. Liu, Y. Fu, G. Shen, A comprehensive survey on traffic missing data imputation, IEEE Trans. Intell. Transp. Syst. 25 (12) (2024) 19252–19275.

[14] D. Salinas, V. Flunkert, J. Gasthaus, T. Januschowski, DeepAR: Probabilistic forecasting with autoregressive recurrent networks, Int. J. Forecast. 36 (3) (2020) 1181–1191.

[15] C. Du, Y. Wang, S. Song, G. Huang, Probabilistic contrastive learning for long-tailed visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 46 (9) (2024) 5890–5904.

[16] L. Zhu, H. Zhao, Z. Zhu, C. Zhang, X. Kong, Multimodal sentiment analysis with unimodal label generation and modality decomposition, Inf. Fusion 116 (2025) 102787.

[17] P. Han, C. Chen, An efficient cross-view image fusion method based on selected state space and hashing for promoting urban perception, Inf. Fusion 115 (2025) 102737.

[18] Y. Lu, W. Wang, R. Bai, S. Zhou, L. Garg, A.K. Bashir, W. Jiang, X. Hu, Hyper-relational interaction modeling in multi-modal trajectory prediction for intelligent connected vehicles in smart cites, Inf. Fusion 114 (2025) 102682.

[19] J. Cao, X. Wang, G. Chen, W. Tu, X. Shen, T. Zhao, J. Chen, Q. Li, Disentangling the hourly dynamics of mixed urban function: A multimodal fusion perspective using dynamic graphs, Inf. Fusion 117 (2025) 102832.

[20] C. Dong, D.B. Clarke, X. Yan, A. Khattak, B. Huang, Multivariate random-parameters zero-inflated negative binomial regression model: An application to estimate crash frequencies at intersections, Accid. Anal. Prev. 70 (2014) 320–329.

[21] C. Caliendo, M. Guida, A. Parisi, A crash-prediction model for multilane roads, Accid. Anal. Prev. 39 (4) (2007) 657–670.

[22] J. Oh, S.P. Washington, D. Nam, Accident prediction model for railway-highway interfaces, Accid. Anal. Prev. 38 (2) (2006) 346–356.

[23] J. De Oña, R.O. Mujalli, F.J. Calvo, Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks, Accid. Anal. Prev. 43 (1) (2011) 402–411.

[24] J. Ma, K.M. Kockelman, P. Damien, A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods, Accid. Anal. Prev. 40 (3) (2008) 964–975.

[25] B. Sharma, V.K. Katiyar, K. Kumar, Traffic accident prediction model using support vector machines with Gaussian kernel, in: Proceedings of Fifth International Conference on Soft Computing for Problem Solving, 2016, pp. 1–10.

[26] Y. Lv, S. Tang, H. Zhao, Real-time highway traffic accident prediction based on the k-nearest neighbor method, in: 2009 International Conference on Measuring Technology and Mechatronics Automation, vol. 3, 2009, pp. 547–550.

[27] J. Abellán, G. López, J. De OñA, Analysis of traffic accident severity using decision rules via decision trees, Expert Syst. Appl. 40 (15) (2013) 6047–6054.

[28] Q. Chen, X. Song, H. Yamada, R. Shibasaki, Learning deep representation from big and heterogeneous data for traffic accident inference, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 30, 2016.

[29] C. Chen, X. Fan, C. Zheng, L. Xiao, M. Cheng, C. Wang, Sdcae: Stack denoising convolutional autoencoder model for accident risk prediction via traffic big data, in: 2018 Sixth International Conference on Advanced Cloud and Big Data, CBD, 2018, pp. 328–333.

[30] H. Ren, Y. Song, J. Wang, Y. Hu, J. Lei, A deep learning approach to the citywide traffic accident risk prediction, in: 2018 21st International Conference on Intelligent Transportation Systems, ITSC, 2018, pp. 3346–3351.

[31] W. Jiang, J. Luo, Graph neural network for traffic forecasting: A survey, Expert Syst. Appl. 207 (2022) 117921.

[32] P. Trirat, S. Yoon, J.G. Lee, MG-TAR: Multi-view graph convolutional networks for traffic accident risk prediction, IEEE Trans. Intell. Transp. Syst. 24 (4) (2023) 3779–3794.

[33] Z. Zhou, Y. Wang, X. Xie, L. Chen, C. Zhu, Foresee urban sparse traffic accidents: A spatiotemporal multi-granularity perspective, IEEE Trans. Knowl. Data Eng. 34 (8) (2022) 3786–3799.

[34] X. Liu, Z. Zhang, L. Lyu, Z. Zhang, S. Xiao, C. Shen, P.S. Yu, Traffic anomaly prediction based on joint static-dynamic spatio-temporal evolutionary learning, IEEE Trans. Knowl. Data Eng. 35 (5) (2023) 5356–5370.

[35] S. Wang, Y. Zhang, X. Piao, X. Lin, Y. Hu, B. Yin, Data-unbalanced traffic accident prediction via adaptive graph and self-supervised learning, Appl. Soft Comput. 157 (2024) 111512.

[36] K. Saleh, A. Grigorev, A.S. Mihaita, Traffic accident risk forecasting using contextual vision transformers, in: 2022 IEEE 25th International Conference on Intelligent Transportation Systems, ITSC, 2022, pp. 2086–2092.

[37] A. Grigorev, K. Saleh, A.S. Mihaita, Traffic accident risk forecasting using contextual vision transformers with static map generation and coarse-fine-coarse transformers, in: 2023 IEEE 26th International Conference on Intelligent Transportation Systems, ITSC, 2023, pp. 4762–4769.

[38] D. Zhuang, S. Wang, H. Koutsopoulos, J. Zhao, Uncertainty quantification of sparse travel demand prediction with spatial-temporal graph neural networks, in: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2022, pp. 4639–4647.

[39] B. Lu, X. Huang, Y. Wu, L. Qian, S. Zhou, D. Niyato, Joint optimization of compression, transmission and computation for cooperative perception aided intelligent vehicular networks, IEEE Trans. Veh. Technol. (2025).

[40] A. Caillot, S. Ouerghi, P. Vasseur, R. Boutteau, Y. Dupuis, Survey on cooperative perception in an automotive context, IEEE Trans. Intell. Transp. Syst. 23 (9) (2022) 14204–14223.

[41] B. Lu, X. Huang, Y. Wu, L. Qian, D. Niyato, C. Xu, Cooperative perception aided digital twin model update and migration in mixed vehicular networks, IEEE Trans. Intell. Transp. Syst. (2024).

[42] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 7132–7141.

[43] Y. Feng, H. You, Z. Zhang, R. Ji, Y. Gao, Hypergraph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 3558–3565.

[44] L. Xia, C. Huang, Y. Xu, J. Zhao, D. Yin, J. Huang, Hypergraph contrastive collaborative filtering, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, 2022, pp. 70–79.

[45] P. Veličković, W. Fedus, W.L. Hamilton, P. Liò, Y. Bengio, R.D. Hjelm, Deep graph infomax, in: International Conference on Learning Representations, vol. 2, 2019, p. 4.

[46] Z. Li, W. Huang, K. Zhao, M. Yang, Y. Gong, M. Chen, Urban region embedding via multi-view contrastive prediction, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, 2024, pp. 8724–8732.

[47] Z. Yu, Y. Lu, Y. Wang, F. Tang, K.C. Wong, X. Li, Zinb-based graph embedding autoencoder for single-cell rna-seq interpretations, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, 2022, pp. 4671–4679.

[48] C. Ma, Y. Zhang, Q. Wang, X. Liu, Point-of-interest recommendation: Exploiting self-attentive autoencoders with neighbor-aware influence, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018, pp. 697–706.

[49] Z. Zhou, Y. Wang, X. Xie, L. Qiao, Y. Li, STUaNet: Understanding uncertainty in spatiotemporal collective human mobility, in: Proceedings of the Web Conference 2021, 2021, pp. 1868–1879.

[50] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. (TIST) 2 (3) (2011) 1–27.

[51] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014, arXiv:1412.3555.

[52] X. SHI, Z. Chen, H. Wang, D.Y. Yeung, W.k. Wong, W.c. WOO, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, in: Advances in Neural Information Processing Systems, 28, Curran Associates, Inc., 2015.

[53] N. Bhardwaj, A. Pal, Bhumika, D. Das, Adaptive context based road accident risk prediction using spatio-temporal deep learning, IEEE Trans. Artif. Intell. 5 (6) (2024) 2872–2883.

[54] M. Chen, H. Yuan, N. Jiang, Z. Bao, S. Wang, Urban traffic accident risk prediction revisited: Regionality, proximity, similarity and sparsity, in: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, 2024, pp. 281–290.

[55] X. Wang, M. Hao, M. Wu, C. Shang, R. Yu, J. Kang, Z. Xiong, Y. Wu, Digital twin-assisted safety control for connected automated vehicles in mixed-autonomy traffics, IEEE Internet Things J. (2024).